Chiang Mai J. Sci. 2010; 37(1) : 21-28
www.science.cmu.ac.th/journal-science/josci.html
Contributed Paper

On Preliminary-prediction Intervals for the Difference Between Two Means with Missing Data

Sa-aat Niwitpong*, **Pawat Paksaranuwat, and Suparat Niwitpong**
Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand.
*Author for correspondence; e-mail: snw@kmutnb.ac.th

Received: 16 March 2009
Accepted: 9 September 2009

ABSTRACT

This paper presents prediction intervals for difference between two future sample means when the distributions are non-normal distributed with missing data. We imputed values for the missing data in a random sample $\{X_{1,i}; i=1,...,m_1\}$ and $\{X_{2,i}; i=1,...,m_2\}$ based on the random hot deck imputation method. The prediction intervals considered are non-adaptive prediction interval (*PI*) and the adaptive interval that incorporates a preliminary test of symmetry for the underlying distributions (PI_{adp}). If the preliminary test fails to reject symmetry for both distributions, the adaptive prediction interval uses the non-adaptive prediction interval; otherwise, it uses the non-adaptive prediction interval to the log-transformed data, then the interval is transformed back. Simulation studies show that under imputation for item nonresponse, the adaptive prediction interval that incorporates the test of symmetry contains the difference between two future sample means better than the non-adaptive prediction interval.

Keywords: coverage probability, imputation, missing data, adaptive prediction interval, preliminary test.

1. INTRODUCTION

The problem of calculating prediction intervals for the difference between the means of two independent normal distributions is an important research topic in reliability and quality control. For instance, in the manufacturing problem case, the researcher may need to use a past sample to predict the difference between two future sample means. Several recent papers have dealt with a prediction interval to contain the difference between two future sample means. A prediction interval to contain the mean of a single future sample was discussed by Hahn [1]. A recent paper by Browne [2] who considered the related problem of evaluating the probability of a future mean difference having the same sign as a past mean difference. Hahn [3] proposed procedures for constructing a prediction interval to contain the difference between two future sample means for two independent normal populations for the case of unequal population standard deviations by applying a confidence interval proposed by Welch [4]. This prediction interval is significant in the sense of its application to a claim of product superiority. Hahn [3] quoted that the application of this

/o

Applied Mathematical Sciences, Vol. 4, 2010, no. 8, 347 - 359

Confidence Interval for the Difference of Two Normal Population Means with a Known Ratio of Variances

Suparat Niwitpong

Department of Applied Statistics, Faculty of Applied Science
King Mongkut's University of Technology North Bangkok, Thailand 10800

Sa-aat Niwitpong

Department of Applied Statistics, Faculty of Applied Science
King Mongkut's University of Technology North Bangkok, Thailand 10800
snw@kmutnb.ac.th

Abstract

Motivated by the recent work of Schechtman and Sherman [The two-sample t-test with a known ratio of variances. Statistical Methodology 4, 508-514, (2007).], we investigate, in this paper, a new exact confidence interval for the difference between two normal population means when the ratio of their variances is known. This is an extension of the case of equal variances where a confidence interval is constructed using an exact t-distribution, as opposed to the case of unequal variances with an *approximate* confidence interval. We derived analytic expressions to find the coverage probabilities and expected lengths of two confidence intervals, the Schechtman-Sherman confidence interval and the Welch-Satterthwaite confidence interval, in comparison with each other. Monte Carlo simulation results indicate that the new confidence interval for the difference between two normal means gives a better coverage probability (and a shorter expected length) than that of the well-known Welch-Satterthewaite confidence interval when a known ratio of their variances is large.

Mathematics Subject Classification: 62F25

Keywords: Coverage probability; Expected length; Welch-Satterthewaite confidence interval

SHORT REPORT
doi: 10.2306/scienceasia1513-1874.2009.35.310

Confidence intervals for the difference between two means with missing data following a preliminary test

Pawat Paksaranuwat, Sa-aat Niwitpong*

Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

*Corresponding author, e-mail: snw@kmutnb.ac.th

Received 2 Feb 2009
Accepted 21 Jun 2009

ABSTRACT: Unit nonresponse and item nonresponse in sample surveys are a typical problem of nonresponse which can be handled by weighting adjustment and imputation methods, respectively. The objective of this study is to compare the efficiency of confidence intervals for the difference between two means when the distributions are non-normal distributed and item nonresponse occurs in the sample. The confidence intervals considered are Welch-Satterthwaite confidence interval and the adaptive interval that incorporates a preliminary test of symmetry for the underlying distributions. The adaptive confidence intervals use the Welch-Satterthwaite confidence interval if the preliminary test fails to reject symmetry for the distributions. Otherwise, the Welch-Satterthwaite confidence interval is applied to the log-transformed data, and then the interval is transformed back. Simulation studies show that the adaptive interval that incorporates the test of symmetry performs better than the Welch-Satterthwaite confidence interval when we imputed values for the missing data in two random samples based on the random hot deck imputation method.

KEYWORDS: coverage probability, imputation, random hot deck method, Welch-Satterthwaite confidence interval

INTRODUCTION

The problem of calculating confidence intervals for the difference between the means of two independent normal distributions is an important research topic in statistics. The common way is to use the Welch-Satterthwaite confidence interval when the population variances are known to be unequal[1]. Miao and Chiou[2] compared three confidence intervals for the difference between two means when both normality and equal variances assumptions may be violated. The confidence intervals considered were the Welch-Satterthwaite interval, the adaptive interval that incorporates a preliminary test (pre-test) of symmetry for the underlying distributions, and the adaptive interval that incorporates the Shapiro-Wilk test for normality as a pre-test. The adaptive confidence intervals use the Welch-Satterthwaite interval if the pre-test fails to reject symmetry (or normality) for both distributions. Otherwise, the Welch-Satterthwaite interval is applied to the log-transformed data and the interval is transformed back. Their study showed that the adaptive interval with a pre-test of symmetry has best coverage among the three intervals considered. The aim of this paper is to generalize Miao and Chiou[2]'s confidence intervals to the missing data case.

Incomplete or missing data in sample surveys generally occurs in two ways: unit nonresponse and item nonresponse[3]. Unit nonresponse occurs if a unit is selected for the sample, but no response is obtained for the unit. Weighting adjustment is often used to handle unit nonresponse. Item nonresponse sometimes occurs for certain questions; either the questions that should be answered are not answered or the answers are deleted during editing. Item nonresponse is usually handled by some form of imputation to fill in missing item values. Brick and Kalton[4] list the main advantages of imputation over other methods for handling missing data. First, imputation permits the creation of a general purpose complete public-use data file with or without identification flags on the imputed values that can be used for standard analyses, such as the calculation of item means (or totals), distribution functions, and quantiles. Secondly, analyses based on the imputed data file are internally consistent. Thirdly, imputation retains all the reported data in multivariate analyses.

As there are a number of imputation methods, it is not immediately clear which method should be chosen, especially when an imputation method may be best in one respect but not in others[5]. Qin et al[6] proposed the random hot deck imputation method to impute the missing values for confidence intervals for the differences between two datasets with missing

Contribution in Mathematics and Applications III
East-West J. of Mathematics, a special volume 2010, pp. 231-252

2-D SOLUTIONS IN RE-ENTRY AERODYNAMICS

Gabriel Mititelu* and Yupaporn Areepong†

**Department of Mathematical Sciences, University of Technology Sydney*
Broadway, NSW 2007, Sydney, Australia
e-mail: Gabriel.Mititelu@uts.edu.au

†*Department of Applied Statistics, Faculty of Applied Science,*
King Mongkut's University of Technology North Bangkok,
Bangkok 10800, Thailand
e-mail:yupaporna@kmutnb.ac.th

Abstract

We derive integrable solutions for the two-dimensional (2-D) re-entry dynamical equations of motion of a space vehicle, under the assumptions of standard atmospheric model. It is desirable to have analytical solutions for this important and practical problem which arise during the atmospheric re-entry phase. Therefore, our solution can be effectively applied to investigate and control the rocket flight characteristics. By setting the initial conditions for the speed, re-entering flight-path angle, altitude, atmosphere density, lift and drag coefficients, the nonlinear differential equations of motion are linearized by a proper choice of the re-entry range angles. By carrying out the closed-form integration, we express the solutions with the Exponential Integral, and Generalized Exponential Integral functions. Theoretical frameworks for proposed solutions as well as, several numerical examples, are presented.

1. Introduction

Since the beginning of the space flight, one of the most important aerodynamic problem encountered in astronautics is the return of satellites and space vehicles

---

* corresponding author
Key words: analytical solutions, Exponential Integral, Generalized Exponential Integral.
2000 AMS Mathematics Subject Classification: 76G25

Contribution in Mathematics and Applications III
East-West J. of Mathematics, a special volume 2010, pp. 253-265

EXPLICIT ANALYTICAL SOLUTIONS FOR THE AVERAGE RUN LENGTH OF CUSUM AND EWMA CHARTS

G. Mititelu*, Y. Areepong†,
S. Sukparungsee† and A. Novikov*

* *Department of Mathematical Sciences, University of Technology Sydney*
Broadway, NSW 2007, Sydney
e-mail: Gabriel.Mititelu@uts.edu.au, Alex.Novikov@uts.edu.au

† *Department of Applied Statistics, Faculty of Applied Science,*
King Mongkut's University of Technology North Bangkok
Bangkok 10800, Thailand
e-mail: yupaporna@kmutnb.ac.th, swns@kmutnb.ac.th

Abstract

We use the Fredholm type integral equations method to derive explicit formulas for the Average Run Length (ARL) in some special cases. In particular, we derive a closed form representation for the ARL of Cumulative Sum (CUSUM) chart when the random observations have hyperexponential distribution. For Exponentially Weighed Moving Average (EWMA) chart we solve the corresponding ARL integral equation when the observations have the Laplace distribution. The explicit formulas obviously takes less computational time than the other methods, e.g. Monte Carlo simulation or numerical integration.

1. Introduction

Cumulative Sum (CUSUM) chart was first proposed by Page (1954) in quality control in order to detect a small shift in the mean of a production process as

Key words: control charts, integral equations, analytical solutions.
2000 AMS Mathematics Subject Classification: 45B05

Songklanakarin J. Sci. Technol.
30 (4), 547-552, Jul. - Aug. 2008

Songklanakarin Journal
of Science and Technology

http://www.sjst.psu.ac.th

Original Article

The variable for the generalized confidence interval for the lognormal mean

Thongkam Maiklad *

Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok, Bang Sue, Bangkok 10800, Thailand.

Received 18 September 2007; Accepted 28 July 2008

Abstract

The purpose of this paper is to propose a new variable for the generalized confidence interval method to estimate the confidence interval of the lognormal mean. In order to evaluate the efficiency of this new method, here called t-generalized method, a simulation study was conducted to examine and compare the coverage probability, interval width, and relative bias of this new method and three other methods, the generalized confidence method of Krishnamoorthy and Mathew, the Modified Cox method, and the Angus's conservative method. The results show that at small sample sizes with large variances, only the t-generalized method and generalized confidence method of Krishnamoorthy and Mathew provide coverage probabilities greater than the nominal level. The t-generalized method is more accurate with a shorter confidence interval than the old generalized confidence method in the case of small sample sizes with large variances.

Keywords: coverage probability, interval width, relative bias, generalized confidence interval, modified cox, Angus's conservative, t-generalized

1. Introduction

The lognormal distribution is a skewed distribution whis is widely used for analyzing the data sets where most of the observations are small, but with a few very large values. Such data, for example, may be the costs of a hospital stay, the incomes of individuals, the height of flood in a river, the amount of Hartmonelly hyaline per gram of soil.

Let X be a random variable having a lognormal distribution, ~ lognormal(μ, σ^2). Then $Y = \log(X)$ has a normal distribution, $N(\mu, \sigma^2)$. The density of X is

$$f(x;\mu,\sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} \exp(-(\ln x-\mu)^2/2\sigma^2)) ,$$
$$0 < x < \infty, 0 < \sigma < \infty .$$

The mean and variance of X are $E(X) = \theta = \exp(\mu + \sigma^2/2)$ and $V(X) = (\exp(\sigma^2)-1)\exp(2\mu+\sigma^2)$, respectively. The most commonly used method for obtaining the confidence limits for θ is the so-called na ve transformation method. This method constructs a confidence interval for $\exp(\mu)$ which is the median of X. The result of this method is tolerably accurate for θ when σ^2 is relatively small, but becomes intolerably inaccurate as σ^2 increases. The accuracy gets worse as the sample size increases. In 1957, Aitchison and Brown (Land, 1972) suggested an approximate confidence interval method or transformation method which should converge to the exact limits only when the sample size becomes infinitely large. Zhou and Gao (1997) compared the coverage probabilities of from the na ve transformation method, the Cox method, the Angus's conservative method and the parametric bootstrap. The simulation results showed that the parametric bootstrap method was the most appropriate method for small variances, whereas Angus's conservative method always gave coverage probabilities more than the nominal level. After Weerahandi (1993) developed the generalized confidence interval, Krishnamoorthy and Mathew (2003) compared the upper limit of the 95% confidence interval of $\ln\theta$ from this method with the Land formula and the parametric bootstrapt method. The result showed

*Corresponding author.
Email address: tks@kmutnb.ac.th

INTERNATIONAL JOURNAL OF INTELLIGENT TECHNOLOGY AND APPLIED STATISTICS
VOL.1, NO.1 (2008) PP.1-33

Statistical Hypothesis Testing Under Interval Uncertainty: An Overview

VLADIK KREINOVICH[1], HUNG T. NGUYEN[2*] & SA-AAT NIWITPONG[3]

[1] *Department of Computer Science, University of Texas, El Paso, TX 79968, USA*
[2] *Department of Mathematical Sciences, New Mexico State University, Las Cruces, NM 88003, USA*
[3] *Department of Applied Statistics, King Mongkut's University of Technology, Bangkok, Thailand*

ABSTRACT

An important part of statistical data analysis is hypothesis testing. For example, we know the probability distribution of the characteristics corresponding to a certain disease, we have the values of the characteristics describing a patient, and we must make a conclusion whether this patient has this disease. Traditional hypothesis testing techniques are based on the assumption that we know the exact values of the characteristic(s) x describing a patient. In practice, the value $\widetilde{x}$ comes from measurements and is, thus, only known with uncertainty: $\widetilde{x} \neq x$. In many practical situations, we only know the upper bound Δ on the (absolute value of the) measurement error $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$. In such situation, after the measurement, the only information that we have about the (unknown) value x of this characteristic is that x belongs to the interval $[\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.

In this paper, we overview different approaches on how to test a hypothesis under such interval uncertainty. This overview is based on a general approach to decision making under interval uncertainty, approach developed by the 2007 Nobelist L. Hurwicz.

Keywords:

1 Formulation of the problem

Statistical hypothesis testing is important. An important part of statistical data analysis is hypothesis testing.

Examples. For example, we know the probability distribution of the characteristics corresponding to a certain disease, we have the values of the characteristics describing a patient, and we must make a conclusion whether this patient has this disease.

Another example is when we want to check whether a newly proposed treatment is effective against a disease. In this case, we have a distribution corresponding to un-treated patients, and we want to check whether the values corresponding to the treated patients fit within the same distribution.

*Corresponding author: hunguyen@nmsu.edu

International Journal of Intelligent Technology and Applied Statistics
Vol.1, No.2 (2008) pp.73-85

Prediction interval on the difference between two future sample means with a known ratio of variances.

Suparat Niwitpong and Sa-aat Niwitpong*

Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok, Thailand 10800

ABSTRACT

Motivated by the recent work of Schechtman and Sherman [(2007). The two-sample t-test with a known ratio of variances. Statistical Methodology 4, 508-514], we investigate, in this paper, a new exact prediction interval on the difference between two future sample means with a known ratio of variances. This is an extension of the case of equal variances where a prediction interval is constructed using an exact t-distribution, as opposed to the case of unequal variances with an *approximate* prediction interval. Asymptotic coverage probability and expected length for our proposed prediction interval are derived.

Monte Carlo simulation results indicate that the new prediction interval for the difference between two future sample means gives a better coverage probability than the existing Hahn prediction interval, as well as a shorter expected length.

Keywords: Known ratio of variance; Prediction interval; Two future sample means

1. Introduction

Prediction interval has played an important role in reliability and quality control. Hahn [6] and Cheung et al. [3] gave many applications in quality control that require the construction of two-sided prediction intervals, one-sided simultaneous prediction intervals and simultaneous prediction intervals for all pairwise differences among the means. The prediction interval for the mean of a single future sample was first discussed by Baker [1], and later by Hahn, see [5] and references therein. Hahn [6] also examined the prediction interval on the difference between two future sample means. Cheung et al. [3] discussed simultaneous prediction intervals, based on [6], for multiple comparisons with a standard. In this paper, we re-examine the prediction interval for the difference between two future sample means of Hahn [6] with a known ratio of variances. Schechtman and Sherman [11] described "a situation when a known ratio of variances arises in practice when two instruments report (averaged) response to the same object based on a different number of replicates.

*Corresponding author: snw@kmutnb.ac.th

EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 3, No. 3, 2010, 435-442
ISSN 1307-5543 – www.ejpam.com

SPECIAL ISSUE ON
GRANGER ECONOMETRICS AND STATISTICAL MODELING
DEDICATED TO THE MEMORY OF PROF. SIR CLIVE W.J. GRANGER

A Generalization of Durbin-Watson Statistic

A. K. Gupta[1]*, D. G. Kabe[2], and S. NiwitPong[3]

[1] *Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, USA*
[2] *5971 Greensboro Dr., Mississauga, Ontario, Canada*
[3] *Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Thailand*

Abstract. Two generalizations of the Durbin-Watson Statistic d, for testing that the serial correlation, in a given univariate normal regression model, is zero, to its multivariate counter part, are proposed. In the univariate case the moments of d are obtained in terms of generalized gamma functions. Our methodology is based on the generalized quadratic form of the central Wishart distribution.

2000 Mathematics Subject Classifications: 62M10,62G10
Key Words and Phrases: Additive outlier; AR(1); Predictor; Prediction interval; Unit toot test

1. Introduction

For the univariate normal linear regression model

$$Y = X\beta + e, e \sim N(0, \sigma^2 I) \tag{1}$$

where Y is an n component (column) vector, β has q components, X is $n \times q$ and of rank $q < n$, σ^2 is unknown, the Durbin-Watson statistic d is defined as follows,

$$\begin{aligned}(Y - X\beta)'(Y - X\beta) &= (\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) + Y'(I - X(X'X)X')Y \\ &= (\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) + Y'QQ'Y,\end{aligned}$$

$\hat{\beta} = (X'X)^{-1}X'Y$ and $Q'Q = I$, Q is $n \times m$, m=(n-q) matric of rank $m < n$. It follows that

$$Y'QQ'Y = f'f, \tag{2}$$

*Corresponding author.

Email address: gupta@bgsu.edu (A. Gupta)

INTERNATIONAL JOURNAL OF INTELLIGENT TECHNOLOGY AND APPLIED STATISTICS
VOL.1, NO.2 (2008) PP.19-41

5

Hypothesis Testing with Interval Data: Case of Regulatory Constraints

Sa-aat Niwitpong[1], Hung T. Nguyen[2]
Vladik Kreinovich[3]* and Ingo Neumann[4]

[1] *Department of Applied Statistics, King Mongkut's University of Technology, North Bangkok, Bangkok 10800, Thailand*
[2] *Department of Mathematical Sciences, New Mexico State University, Las Cruces, NM 88003, USA*
[3] *Department of Computer Science, University of Texas at El Paso, El Paso, TX 79968, USA*
[4] *Geodetic Institute, Leibniz University of Hannover D-30167 Hannover, Germany*

ABSTRACT

In many practical situations, there exist regulatory thresholds: e.g., a concentration of certain chemicals in the car exhaust cannot exceed a certain level, etc. In this paper, we describe how to make accept/reject decisions under measurement or expert uncertainty in case of regulatory and expert-based thresholds – where the threshold does not come from a detailed statistical analysis.

This paper expands our conference paper [23].

Keywords: Hypothesis testing, Interval data; Regulatory constraints

1. Hypothesis testing: a general problem

In many practical situations, it is desirable to check whether a given object (or situation) satisfies a given property. For example, we may want to check whether a patient has flu, whether a building or a bridge is structurally stable, etc.

In statistics, this problem is called *hypothesis testing:* we have a hypothesis – that a patient is healthy, that a building is structurally stable – and we want to test this hypothesis based on the available data. This hypothesis is usually called a *null hypothesis*, meaning that:

- if this hypothesis is satisfied then no ("null") action is required,
- while if this hypothesis is not satisfied, then we need to undertake some action: cure a patient, reinforce (or even evacuate) the structurally unstable building, etc.

*Corresponding author: vladik@utep.edu

15

INTERNATIONAL JOURNAL OF INTELLIGENT TECHNOLOGIES AND APPLIED STATISTICS
VOL.3, NO.3 (2010) PP.335-352

Statistical Estimation of Asset Pricing in Case of Heavy-Tailed Distributions

Wararit Panichkitkosolkul* and Sa-aat Niwitpong

Department of Applied Statistics, Faculty of Applied Science
King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

ABSTRACT

Under the classical assumption that the second moment of an option payoff variable is finite, the empirical estimator of option pricing has already been well studied in the literature. However, the result is not applicable when the option payoff variable follows any distribution with infinite second moment, which is a frequent situation in practice. In this paper we propose a new estimator and confidence interval of a contingent claim pricing which is applicable in the case of heavy-tailed distributions.

Keywords: Asset pricing; Option pricing; Wang's distortion function; Heavy-tailed distribution; Hill estimator; Weissman estimator

1. Introduction

In a stock market, an option is a contract between a buyer and a seller that gives the buyer of the option the right, but not the obligation, to buy a specified stock on the option's expiration date denoted by T, at an agreed price, the strike price denoted by K. An option has the objective to transfer a risk from one part to another, against a specific payment.

The value of an option at time T is called the payoff of an option which is

$$X(T) = (S(T) - K)^+ = \max\{S(T) - K, 0\},$$

where $S(T)$ is the price of the stock at time T. It is so because, if $S(T) > K$, then the owner of the option only needs to pay K for something worth more. He can then exercise his exercise his right, buy y stocks at price K each (if he has paid for

* Corresponding author: wararit@mathstat.sci.tu.ac.th

Journal of Uncertain Systems
Vol.4, No.4, pp.296-300, 2010
Online at: www.jus.org.uk

16

From Abstract Natural Numbers to Physical Natural Numbers: A Probabilistic Approach

Wararit Panichkitusolkul, Supaporn Nontanum,
Sa-aat Niwitpong*, Wichai Surachendkiati
[1] *Department of Applied Statistics, Faculty of Applied Science,*
King Mongkut's University of Technology North Bangkok,
1518 Pibulsongkram Rd., Bangsue, Bangkok 10800 Thailand

Received 12 January, 2009; Revised 30 January, 2010

Abstract

Natural numbers originated as a way to describe the result of counting procedures. In quantum physics, the results of counting are probabilistic, so, in general, real counting leads to a random natural number – a probability distribution on the set of all natural numbers. From the practical viewpoint, events with a very small probability (smaller than some threshold ε) cannot occur. Therefore, it is reasonable say that a random natural number represents an integer n if the probability of n is $> \varepsilon$, while the probability of every other number is $\leq \varepsilon$. For thus defined physical natural numbers, we analyze how their properties differ from the properties of the standard mathematical natural numbers. Specifically, we analyze the following natural question: if a represents n and b represents m, what are the possible representations for $a + b$?

Keywords: integers, quantum physics, probabilistic uncertainty

1 Formulation of the Problem

Not all natural numbers are physically meaningful. Natural numbers originated from the need to count real objects. Reasonably small natural numbers can indeed be interpreted as the corresponding numbers of objects. However, very large abstract integers, integers like $10^{10^{10}}$ which are larger than the number of particles in the Universe, cannot be thus represented. A natural question is: what will happen if we only allow physically meaningful natural numbers? This question was analyzed in the past from the philosophical and logical viewpoint; see, e.g., [1, 2, 3, 9, 10].

In this paper, we analyze the same question from a more physical viewpoint; in this analysis, we follow ideas from [4, 5, 6, 7, 8].

2 Towards a Definition of a "Physical" Natural Number

Towards a definition of a physical natural number. It is reasonable to identify, e.g., number 1 with situations in which the result of a counting procedure can be 1 but cannot be anything else.

Real physical natural numbers are probabilistic. The formalization of the above idea is complicated by the fact that according to quantum physics, all predictions are probabilistic.

In particular, for every physical counting procedure applied to a physical state, the result is, in general, a *random* natural number – in other words, a probability distribution on the set of all natural numbers in which can get different values i with different probabilities $p(i) \geq 0$ ($\sum_i p(i) = 1$).

In these terms, how can we describe the idea that some values are possible and some are not?

*Corresponding author. Email: snw@kmutnb.ac.th (S. Niwitpong)

International Journal of Innovative Management, Information & Production
Volume 1, Number 1, September 2010
ISME International ©2010 ISSN 2185-5439
PP. 49-55

MULTISTEP-AHEAD PREDICTORS FOR A GAUSSIAN AR(1) PROCESS WITH ADDITIVE OUTLIERS FOLLOWING THE UNIT ROOT TESTS

WARARIT PANICHKITKOSOLKUL AND SA-AAT NIWITPONG
Department of Applied Statistics
King Mongkut's University of Technology North Bangkok
Bangkok, 10800, Thailand
wararit@mathstat.sci.tu.ac.th
snw@kmutnb.ac.th

ABSTRACT. *Recent works of Diebold and Kilian (2000) and Niwitpong (2007, 2009) indicated that a one-step-ahead predictor for an AR(1) process can be improved by using the preliminary unit root tests. This paper extends these mentioned concepts to the multistep-ahead predictors of a Gaussian AR(1) process with additive outliers. The following predictors are considered: the standard predictor, the predictor following the Dickey-Fuller unit root test, and the predictor following the Shin et al. (1996) unit root test. The relative efficiencies of all predictors based on the prediction mean square error are compared through simulation studies. Simulation results have shown that the unit root test can improve the preciseness of the multistep-ahead predictors for a near non-stationary AR(1) process with additive outliers.*

Keywords: AR(1) Process; Additive Outliers; Predictor; Unit Root Test

1. Introduction. Outliers, or aberrant observations, in a time series can have adverse effects on parameter estimation including prediction. Fox (1972) proposed two parametric models for studying outliers in a time series. He defined additive outliers (AO) and innovations outliers (IO). In this paper, we focus solely on the additive outliers because these outliers are more hurtful than innovations outliers (2001). Let $\{X_t ; t = 2,3,\dots,n\}$ be the first-order autoregressive process, AR(1), satisfying

$$X_t = \mu + \rho(X_{t-1} - \mu) + e_t, \quad (1)$$

where μ is the mean of the process, ρ is an autoregressive parameter, $\rho \in (-1,1)$, and e_t is a sequence of independent and identically distributed $N(0,\sigma_e^2)$ random variables. If $\rho = 1$, then the model (1) is called the random walk model and hence non-stationary. The random walk model is given by

$$X_t = X_{t-1} + e_t. \quad (2)$$

However, if $|\rho| < 1$, then it can be shown that the process is stationary. For a near non-stationary process, i.e. $|\rho| \to 1$, the mean, variance and autocorrelation function of this process are not constant through time. An observed time series Y_t has an additive outlier at time T of size δ if it satisfies $Y_t = X_t + \delta I_t^{(T)}$, where $I_t^{(T)}$ is an indicator variable such that $I_t^{(T)} = 1$ if $t = T$, and $I_t^{(T)} = 0$ if $t \neq T$.

(11)

Journal of Management Science and Statistical Decision

Vol. 6 No. 4, December 2009

Contents

Financial Risk Modeling: Probabilistic Analysis and Statistics 1
Hung T. Nguyen

DEA-Based Evaluation of China's Recycling Economy Efficiency: Case Study of First Groupof Pilot Provinces and Cities 9
Li Tong, Chen Liping

Research of Incentive Mechanism for Developing of Human Resource of Marine Science and Technology 14
Cui Wanglai, Chen Bing

The Exploring on Construction Project Target Control 19
Dongliang Yuan, Delong Xu

The Impact of Capital Structure and Debt Maturity Structure on SMEs' Performance 23
Yang Yimin, Yang Xianting

Dynamical Fuzzy Decision Model of Agricultural Product Logistics Scheme Selection 27
Lingyun Zhou, Yanru Zhu, Gang Zhao, Jianfeng Luo

An Extended Study of the "HU Theory" 31
Hu Zuguang, Hu Jing

Assess the Goodness of Fit for Risk Aversion Parameter of First Price Auction via Nonparametric Method 38
Xin An, Shulin Liu, Shuo Xu

Prediction Intervals for an Unknown Mean Gaussian AR(1) Process Following Unit Root Tests 43
Sa-aat Niwitpong, Wararit Panichkitkosolkul

Vol.6 No4.(2009) p43-51

Prediction Intervals for an Unknown Mean Gaussian AR(1) Process Following Unit Root Tests

Sa-aat Niwitpong
Department of Applied Statistics
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
snw@kmutnb.ac.th

Wararit Panichkitkosolkul
Department of Applied Statistics
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
wararit@mathstat.sci.tu.ac.th

***Abstract*—This paper presents two new one-step-ahead prediction intervals for an unknown mean Gaussian AR(1) process. We propose the simple prediction interval based on the residuals model, PI_a, and the prediction interval following the unit root tests, PI_f. The unit root tests applied in this paper are the Dickey-Fuller unit root test, the Phillips-Perron unit root test, the weighted symmetric unit root test, and the Elliott-Rothenberg-Stock unit root test. The coverage probabilities of all prediction intervals are derived. The performance of the proposed prediction intervals is assessed through Monte Carlo simulation studies. Simulation results have shown that all prediction intervals have minimum coverage probabilities 0.95 for all the autoregressive parameter values. Moreover, the expected lengths of prediction intervals PI_f are shorter than that of a prediction interval PI_a when the autoregressive parameter value is close to one.**

Keywords-AR(1); Unit Root Test; Prediction Interval; Preliminary test; Residual Model

I. INTRODUCTION

Recently, there has been increasing interest in constructing prediction intervals for an autoregressive process, see for example the standard textbooks of Box et al. [1] and Wei [2]. The conditional prediction interval for an autoregressive moving average model was proposed by de Luna [3]. However, this prediction interval does not has a minimum coverage probability at the nominal confidence level for small and moderate sample sizes. Halkos and Kevork [4] proposed a prediction interval for a stationary AR(1) process with an almost unit root by considering the prediction interval based on the random walk model when the autoregressive parameter, ρ, is close to one. They found that their proposed prediction interval has less coverage probability than the nominal confidence level when ρ is close to one. Several authors also used the bootstrap methods for calculating prediction intervals, see for example Kim ([5], [6], [7]), Alonso et al. ([8], [9]) and Clements and Kim [10] and references cited therein. Although prediction intervals based on bootstrap methods are easy to construct, some of these prediction intervals have coverage probabilities less than the nominal level $1-\alpha$, see the simulation results, Table 1 of Alonso et al. [8] and Table 2 of Clements and Kim [10]. In this paper, we emphasis constructing the simple prediction intervals, based on the residual model described by Olive [11] for an AR(1) process with an almost unit root and these prediction intervals have minimum coverage probabilities $1-\alpha$.

We consider an AR(1) $\{Y_t; t=1,2,3,...,T\}$ which satisfies

$$Y_t = \mu + \rho(Y_{t-1}-\mu)+e_t \qquad (1)$$

where μ is the population mean, ρ is an autoregressive parameter, $\rho\in(-1,1)$, and $e_t \sim N(0,\sigma_e^2)$. For $\rho=1$, model (1) is called the random walk model, otherwise; it is called a stationary AR(1) process when $\rho<1$. In this paper, it is therefore of interest to construct the prediction interval for (1) when there is an uncertainty of this process, i.e. ρ is close to one. Applications in econometrics of (1) in econometrics, when it is doubtful whether this process is a stationary process or a random walk process, have been described by Hamilton ([12], pp. 501-503). Hamilton described the need to use the unit root hypothesis test to find the correct model for the series of the nominal interest rate of the United States from 1947-1949 and the real GNP for the United States from 1947-1989, see Figures 17.2-17.3 of Hamilton ([12], pp. 503). Hamilton also described that there is no guarantee in economic theory suggesting that the nominal interest rate series should be a deterministic time trend model, although Figure 17.2 shows an upward trend over the sample data. The model for these data might be a random walk without trend or a stationary process model with a constant term. Therefore, the interesting question arises whether a one-step-ahead prediction interval of the true model of the nominal interest rate series is computed from a random walk model or a stationary model. To answer this question, the unit root test, see e.g. Dickey and Fuller [13], will be used to choose between these models. If the unit root test does not reject the null hypothesis; $H_0:\rho=1$ against the alternative hypothesis; $H_a:\rho<1$, we conclude that this series is a random walk model. We proceed to construct the prediction interval for an AR(1) process using the random walk model. However, if the hypothesis $H_0:\rho=1$ is rejected, the prediction